Best Available Copy

ESD-TR-78-203

LEVEL ~~III~~

B028 521

12

AD A060714

DDC FILE COPY

Semiannual Technical Summary

B028 561

DDC
NOV 2 1978
F

Information Processing Techniques Program

Volume II: Wideband Integrated Voice/Data Technology

31 March 1978

Prepared for the Defense Advanced Research Projects Agency
under Electronic Systems Division Contract F19628-78-C-0002 by

Lincoln Laboratory

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

LEXINGTON, MASSACHUSETTS

Approved for public release; distribution unlimited.

78 10 23 145

The work reported in this document was performed at Lincoln Laboratory, a center for research operated by Massachusetts Institute of Technology. This work was sponsored by the Defence Advanced Research Projects Agency under Air Force Contract F19628-78-C-0002 (ARPA Order 2929).

This technical report has been reviewed and is approved for publication.

FOR THE COMMANDER

Raymond L. Loiselle

Raymond L. Loiselle, Lt. Col., USAF
Chief, ESD Lincoln Laboratory Project Office

(12)

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

LINCOLN LABORATORY

INFORMATION PROCESSING TECHNIQUES PROGRAM
VOLUME II:
WIDEBAND INTEGRATED VOICE/DATA TECHNOLOGY

SEMIANNUAL TECHNICAL SUMMARY REPORT
TO THE
DEFENSE ADVANCED RESEARCH PROJECTS AGENCY

1 OCTOBER 1977 – 31 MARCH 1978

ISSUED 31 AUGUST 1978

Approved for public release; distribution unlimited.

LEXINGTON MASSACHUSETTS

ABSTRACT

This report describes work performed on the Wideband Integrated Voice/Data Technology program sponsored by the Information Processing Techniques Office of the Defense Advanced Research Projects Agency during the period 1 October 1977 through 31 March 1978.

CONTENTS

	Abstract	iii
	Introduction and Summary	vii
I.	SPEECH CONCENTRATION REQUIREMENTS STUDY	1
	A. Introduction	1
	B. Elements of Speech Concentration Facility	1
	C. Design Objectives	2
	D. Access Area Structures	4
	1. Centralized (Star) Geometry	5
	2. Ring Geometry	6
	3. ETHERNET	7
	4. Local Voice Network	8
	E. Voice-Terminal Definition	14
	1. Speech Processor	15
	2. Privacy Module	16
	3. Protocol Processor	17
	4. Modem	18
	F. Concentrator Functions	19
II.	ADAPTIVE VARIABLE-RATE PACKET SPEECH NETWORKING	21
	A. Introduction	21
	B. Network Configuration	22
	C. Feedback Schemes	23
	D. Experiments	24
APPENDIX	– Synchronization Issues in Packet Speech Communication	27

INTRODUCTION AND SUMMARY

This report is the first Semiannual Technical Summary for the DARPA-sponsored Wideband Integrated Voice/Data Technology Program. The goal of this program is the investigation and development of techniques for integrated voice and data communication in packetized networks which include wideband common-user satellite links. Specific areas of concern are the concentration of statistically fluctuating volumes of voice traffic; the adaptation of communication strategies to conditions of jamming, fading, and traffic volume; and the eventual interconnecting of wideband satellite networks to terrestrial systems.

The technology background for this program is provided by past developments in the DARPA-sponsored Packet Speech Program and Communications Adaptive Internetting Program. The Packet Speech program will continue to develop basic supporting technology in the area of digitized voice communications.

Plans call for the establishment of an experimental wideband satellite network to serve as a unique facility for the realistic investigation of voice/data networking strategies. This facility will be jointly sponsored by DARPA and DCA and will include four ground stations sharing a leased domestic wideband satellite transponder.

The current report covers work in two areas: a study of speech concentration requirements and a simulation of a technique for adaptive variable-rate packet speech networking. The speech concentration requirements study begins with an identification of the basic elements of a speech concentration facility and an outline of design objectives relating to the separation of functions among these elements. Access area options, voice terminal design issues, and concentrator requirements are then each discussed in more detail. The adaptive-networking efforts are based on an embedded speech coding technique coupled with priority-oriented packet handling and end-to-end flow control. The effects of these strategies are being studied in the context of a simple network topology consisting of a central node through which pass 16 paths connecting 4 nodes on either side.

INFORMATION PROCESSING TECHNIQUES PROGRAM

WIDEBAND INTEGRATED VOICE/DATA TECHNOLOGY

I. SPEECH CONCENTRATION REQUIREMENTS STUDY

A. Introduction

The experimental wideband network represents the first opportunity for packet speech experiments in which a large number of simultaneously active voice users can be accommodated in an integrated voice/data network environment. In contrast to the ARPANET and Atlantic Packet Satellite Network environments, where the functions required for interfacing a few speech processors to the network could be accomplished in standard host minicomputers, voice experiments on the wideband net will require speech concentration facilities capable of providing access for numbers of voice terminals to individual network nodes. These concentrators will initially be connected to high-capacity SIMPs (Satellite Interface Message Processors) to support a variety of important speech communication experiments on the satellite channel. In the longer term, it is anticipated that similar concentration facilities will supply the voice terminal access and voice traffic regulation functions in a combined terrestrial/satellite wideband system.

Lincoln has initiated a study aimed at defining speech concentration requirements and making recommendations regarding the functional capabilities and architectural design of speech concentration systems. Issues to be addressed include: the separation of functions between speech terminals and concentrators, the structure of the access area, the role of traffic emulation modules in early experiments, speech traffic flow control mechanisms, and compatibility with the variety of network switch types which might be included in the terrestrial/satellite network.

This first report on the study begins with an identification of the basic elements of a speech concentration facility. Broad design objectives relating to the separation of functions among these elements are outlined. Access-area options, voice-terminal design issues, and concentrator requirements are then each discussed in more detail.

B. Elements of Speech Concentration Facility

A general structure for a speech concentration facility is shown in Fig. 1. Three essential components are needed: (1) the individual voice terminals at user locations, (2) an access area which provides communication between the terminals and a central facility, and (3) a concentrator which provides multiplexing/demultiplexing and other necessary functions to interface the local voice terminal community to the wideband network. In the experimental program, a traffic emulation module will also be required so that the network can be tested with substantial voice traffic loads without the need for initially activating a large community of voice users.

The purpose of this study is to set forth and compare alternatives for access-area designs, voice-terminal configurations, and concentrator characteristics, and to specify partitioning options among the functions of the three main system elements. These partitioning options are bounded at one extreme by incorporating all the speech and networking functions in a very flexible and perhaps remotely programmable terminal that essentially acts as a combination voice processor and network host. At the other extreme, a majority of the networking tasks such as packetization, dial-up and conferencing protocols, packet reconstitution algorithms, etc., are

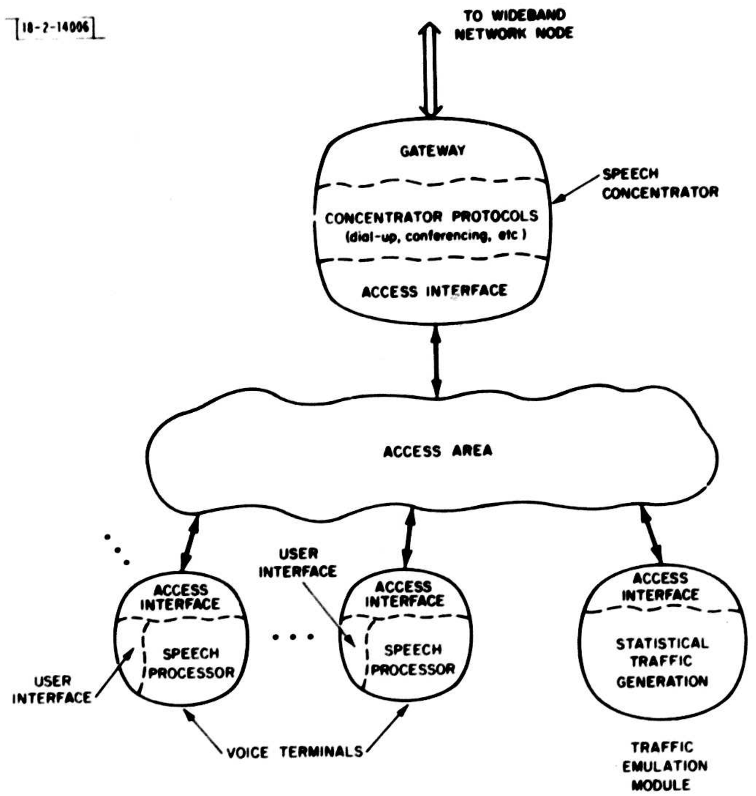

Fig. 1. System elements and functional separation.

effected in the concentrator, with the terminals supporting only those speech-related functions that absolutely have to be performed at the user site. Although we anticipate that continued progress in digital large scale integration will eventually turn the combined voice-processor/ network-host terminal into an economically viable option, we do not feel that it represents a reasonable choice for the large number of terminals that will be needed for relatively-near-term experiments in the wideband testbed system. We have therefore focused our initial efforts on functional partitionings that offer extreme simplicity of voice-terminal design while preserving the system flexibility that is critical to the experimental environment. Our selection of system options has been guided by some general design objectives for multi-user packet speech, which will be outlined in the next section. Within the guidelines of these design objectives, cost-effectiveness and flexibility are the key criteria for distinguishing among alternatives.

C. Design Objectives

General considerations regarding the requirements for multi-user packet speech have led to a set of design objectives for speech terminals and concentration systems for use in the wideband experimental network. The objective is the desire for a separation of functions between system elements which will allow maximum flexibility and growth potential. The desire is to provide a framework within which a variety of speech terminal types can gain access to the network facility.

Design objectives which have been identified based on broad systems issues are listed below. Additional considerations will evolve as we investigate hardware implications and protocol requirements in greater detail.

(1) Speech terminals should be independent of network characteristics or protocols.

(2) Network switches should not be required to have knowledge of speech algorithms or data formats.

(3) Concentrators should not be required to perform speech-algorithm-related functions, so that (software or hardware) changes in the concentrator will not be necessary each time a new speech algorithm is introduced. For example, silence detection and the reconstruction of silence intervals of proper duration are speech terminal issues and should ideally be performed in the terminals.

(4) Dial-up and conferencing protocols are networking issues and should be dealt with in concentrators. The associated control communication between terminals and concentrators should be carried out via simple local protocols involving touch-tone-like user interface devices at the terminals.

(5) Transformations of the data for privacy purposes should be possible at the individual voice terminal level, independent of wideband bulk encryption which may be carried out at the concentrator/network interface. This reinforces objective (3) above, since the provision of speech-algorithm-related functions in the concentrator becomes infeasible when the concentrator has access to the voice stream only in scrambled form.

(6) Network-specific packetization and transmission functions should reside in the concentrators in the form of gateway-like software (see Fig. 1) which can be adapted to a variety of networks.

(7) The specific details of the voice access area design should not be reflected in or influenced by network protocol requirements. Such details should be a private issue between a concentrator and its local voice terminal community. The major function of the access-area design is to efficiently and economically provide voice-terminal connectivity to and from the concentrator, and to support the concentrator's packetization/depacketization and multiplexing/demultiplexing roles. Both the terminals and the concentrator should have simple and separable modules for access-area interfacing, as indicated in Fig. 1.

(8) The introduction of new terminals or the relocation of previously connected terminals should be as simple and convenient as possible.

(9) A traffic emulation module for experimental use should fit gracefully into the access area/concentrator system structure without unduly perturbing the design of that structure.

In the process of attempting to define the above objectives, discussions were carried out with other participants in the ARPA packet speech community. In particular, it was found that independent work at Information Sciences Institute on the issues of interfacing voice terminals to packet networks* has resulted in a similar and generally compatible set of design objectives. The work at ISI has been concerned primarily with concentrator-to-concentrator protocol design issues, whereas Lincoln's emphasis is in the design of access-area techniques and terminal/concentrator interfacing. In addition, discussions with Bolt, Beranek and Newman (BBN) regarding ongoing work on block-oriented privacy techniques† for packetized systems via a BCR (Black-Crypto-Red) approach have provided valuable inputs regarding the privacy issue for packetized voice.

The following sections describe access-area designs, voice-terminal configurations, and concentrator characteristics that have been considered based on the design objectives outlined above. A driving motivation in the choice of a joint terminal/access area/concentrator design is that of overall system economy. The partitioning of system functions between many small voice terminals and a single large concentrator is critical in view of the fact that large numbers of voice terminals will eventually be deployed. The topology of the access area will influence the choice of that functional partition, and also has major implications with respect to system flexibility and hardware complexity.

D. Access Area Structures

Two generic topologies have been considered for possible access-area use; namely, centralized and distributed. Although radio connectivity within an access area might be appropriate in some special applications, the requirements of the ARPA/DCA wideband integrated network tested are probably best met via the use of direct ohmic connections between the central concentrators and their local communities of voice terminals. Our model has been that of a single concentrator located at a facility such as Lincoln Laboratory, Defense Communications Engineering Center, or ISI, serving a relatively large number of digital voice terminals dispersed throughout an area local to that facility. In addition, a traffic emulation module, capable of producing digital data that simulates the presence of many voice terminals, and requiring a wideband connection to the concentrator, is assumed to exist at each facility.

In the centralized access area configuration, the speech concentrator is independently connected to each voice terminal via separate cables. Distributed geometries include serial (ring-like) arrangements and parallel (ETHERNET-like) organizations of terminals within an access area. These schemes are described in detail below, and they are reviewed in the context of the requirements of the wideband experimental network. The star geometry is rejected for this application on the basis of flexibility limitations and hardware considerations. Ring structures could present reliability problems, but these can probably be overcome. The ETHERNET architecture has attractive features, but it is probably better suited for interactive data traffic than for the steadier traffic flows characteristic of voice. A modified cable network, similar to ETHERNET in geometry, but better matched to the voice terminal/speech concentrator communications environment, is proposed and described in detail.

* R. Cole and D. Cohen, "Issues in Packet Voice Interfacing," Network Speech Compression (NSC) Note No. 123 (February 1978).

† S. T. Walker, "ARPA Network Security Project," EASCON '77, p. 14-5A.

1. Centralized (Star) Geometry

The geometry shown in Fig. 2 is perhaps the simplest from a structural point of view. Each voice terminal is independently connected to the central concentrator via a dedicated cable. The

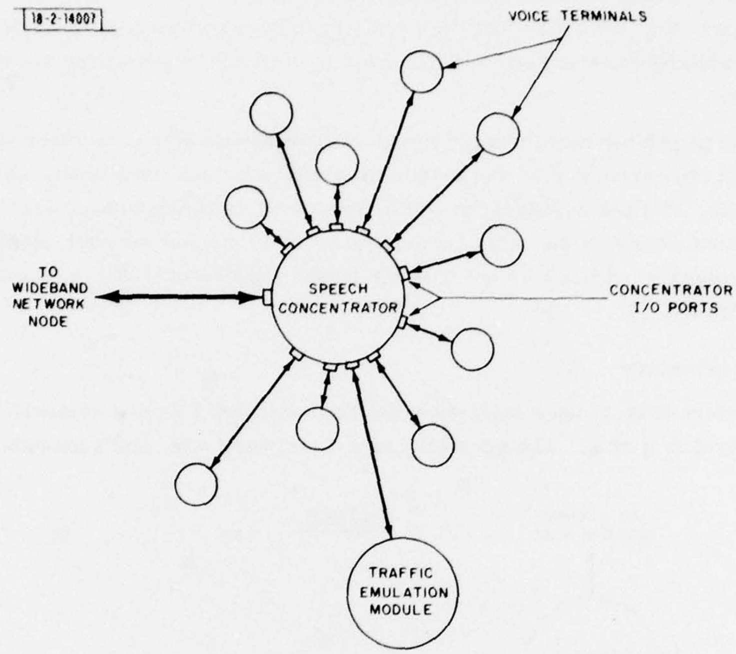

Fig. 2. "Star" (centralized) access-area geometry.

concentrator functions as a multiple I/O controller, dealing with each voice terminal in accordance with conventional priority-based/interrupt-driven I/O handling methods. Data transfers between concentrators and terminals include voice parcels destined for or coming from remote concentrators in the network, voice parcels directed to or coming from other terminals in the same local access area, and private control transactions between terminals and their own concentrators. The latter include dialing information and call status (ringing, busy, etc.) signals that are used during the establishment of a connection, as well as control messages that are required during an ongoing call. Examples include conference control signaling (vote-taking, queue-to-talk, etc.) or vocoder rate-change messages (in adaptive variable rate experiments). The following observations can be made with regard to this access-area configuration:

(a) A separate I/O port is required at the concentrator for each voice terminal. This presents a practical limit on the total number of terminals that might be deployed, even if only a few of them are assumed to be active (off-hook) at a given time.

(b) A wideband port, probably of different design than those used for the individual terminals, will be needed for a traffic emulation module. Thus, although the emulation requirement is a temporary one, it influences the basic design of the concentrator subsystem.

(c) Data transfers between a terminal and its concentrator will require a formatting protocol that allows for control communications as well as for voice data flow. Thus, although a separate wire path exists between each terminal and the concentrator, both devices will have to identify frame or packet boundaries and decode selected portions of the data stream. It is not clear that the resultant logical complexity would be substantially different from that needed in distributed-geometry access areas.

(d) The required bit rate for each terminal-to-concentrator connection link should be determined by the maximum anticipated voice communications bit rate. This then allows for variable as well as fixed-rate protocols in future experiments. Since a separate link is needed for each terminal regardless of whether or not it is off-hook, significant wiring costs are anticipated.

2. Ring Geometry

The ring structure (Fig. 3) uses point-to-point transmission between adjacent terminals conceptually arranged in a ring. The geometry is a distributed one, and succeeds in avoiding

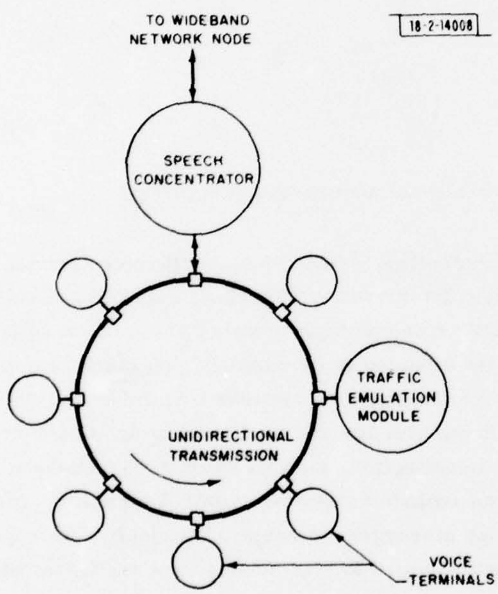

Fig. 3. Ring topology.

some of the difficulties of the above-described centralized architecture. For example, the concentrator requires only a single ring interface, regardless of the number of voice terminals in the system.

An interface at each terminal regenerates its received messages and passes them on to the next terminal in the ring. To transmit, a terminal awaits the receipt of a "control token" bit pattern and then breaks the repeater connection across the interface, gating its message, bit serially onto the ring. The concentrator then copies the message as it passes through its own ring interface. The same process is used for communicating from the concentrator to the terminals.

With the exception of the unidirectional transmission characteristics and its data-regenerating interfaces, the ring structure can be viewed as a shared broadcast medium with a slotted burst time-division transmission protocol. As such, it can be controlled via any of several appropriately selected strategies. For example, the allocation of transmission time slots might be placed under the control of the concentrator. The latter would respond to "off-hook" indications from terminals desiring access to the system, and distribute transmission slots, rate allocations, etc., accordingly.

Although the ring architecture offers some advantages compared to the star topology, the basic concept appears to be weak in the context of overall system reliability. For instance, only a small fraction of telephones in a given population can be expected to be "off-hook" simultaneously. The ring requires that all terminals, both active and inactive, participate in the data regeneration and retransmission process. This raises a serious reliability issue, since the ring interfaces of all terminals are in series. Even if inactive terminals were to be electrically removed from the ring, the failure of a single active terminal could result in an overall system crash. An additional problem in this system is that of introducing new terminals. One basically has to break the ring in order to add a terminal, and this could result in the occasional suspension of system operation. While this might be tolerable in an experimental test bed, it could preclude consideration of the ring as a model for future operational access area designs. We note that other distributed access area configurations might in fact be subject to similar difficulties.

3. ETHERNET

The ETHERNET is a distributed data communications concept developed by Metcalf and Boggs* of the Xerox Palo Alto Research Center (PARC). Closely related variations of the basic notion include the CHAOSNET at M.I.T., and the FARBERNET at the University of California at Irvine. In contrast to the serial ring architecture, the ETHERNET provides a parallel form of connectivity between a community of terminals. The structure, shown in the context of a voice access area in Fig. 4, uses a single coaxial cable as a transmission/reception medium. The concentrator and the voice terminals constantly monitor all the messages which are broadcast over the cable, and each device extracts only those messages that are addressed to it.

The connection of a device to the cable is a passive one, and during their silent periods individual terminals present little or no load to the cable. The presence of a terminal is thus invisible to the system unless it transmits. An ALOHA-like transmission protocol is used in conjunction with collision-sensing hardware in the terminals. In brief, a terminal can "send" when it perceives the cable to be "quiet." If more than one terminal decides to transmit simultaneously, each will sense the presence of the other; both will try again after random waiting periods. The system has the potential for high efficiency since collisions are sensed and terminated before significant amounts of data have been transmitted.

---

* R. M. Metcalfe and D. R. Boggs, "ETHERNET: Distributed Packet Switching for Local Computer Networks," Commun. ACM 19, 395 (1976).

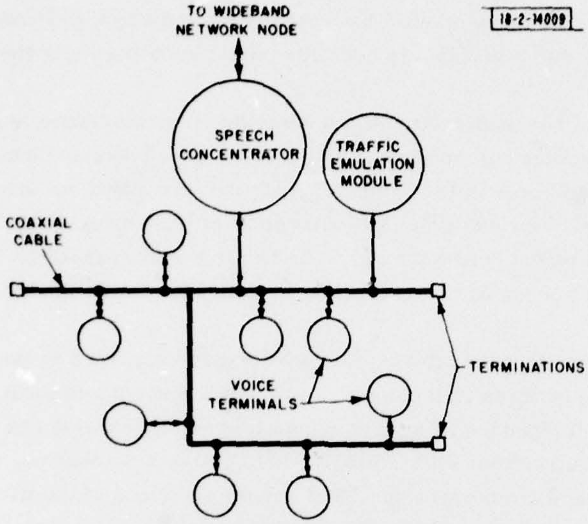

Fig. 4. ETHERNET concept.

A major source of concern in an ETHERNET-like voice access area is the possible mismatch between a random contention-based transmission protocol and the inherently periodic nature of speech traffic. A reasonable model for the output of a voice terminal is that of a periodic sequence of burst transmissions, at least during non-silent intervals. Although burst repetition sizes and rates may vary between the terminals in a given access area, one still expects a greater degree of correlation between packet collisions in this situation than in a system that handles purely random Poisson-distributed data traffic. A possible manifestation of this effect could be that a terminal that encounters collision problems in sending one burst, might be likely to experience the same difficulty in transmitting its next burst.

A second problem area relates to the fact that one device, the concentrator, consumes fully 50 percent of the total system bandwidth utilization any time. This follows from the fact that voice conversations are two way, so that on the average every voice terminal sends and receives (to and from the concentrator) the same amount of traffic. Despite the rapid collision recovery feature of the ETHERNET, one imagines that lockups or other difficulties might arise if many separate terminals attempt to transmit simultaneously. This can happen when the concentrator releases the channel after having captured it for a long uninterrupted period.

Although the above-described problem areas can probably be dealt with via appropriate protocol designs, one suspects that smaller and less expensive terminal configurations will result from a system design that avoids complicated collision-recovery logic and carrier sensing and collision detection hardware. The Local Voice Network described in the next section evolved from consideration of these issues. Its topology is basically that of a modified ETHERNET in which the concentrator is provided with its own private transmission channel.

4. Local Voice Network

a. General Description

The Local Voice Network, Fig. 5, uses separate channels for terminal-to-concentrator and concentrator-to-terminal communications. The two links may be implemented as separate

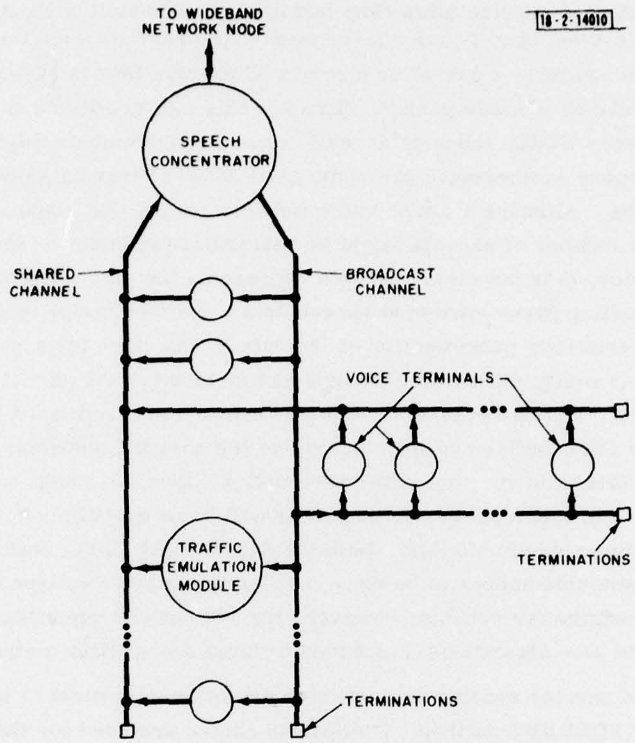

Fig. 5. Local voice network.

coaxial lines or as two distinct frequency bands or time slots sharing the same cable. The main feature of the separation is that the concentrator can send messages to individual voice terminals without the danger of possible collisions due to contention. Terminals are connected to the system such that the presence of a terminal is "invisible" except when it is transmitting. Terminal functions include speech activity detection and silence reconstruction.

Data flow from the concentrator to the terminals is via a packet broadcast transmission protocol that includes a synchronization header, a terminal address, call status information, etc. No restrictions are placed on the contents of the data portions of the packets or bursts other than those imposed by the speech terminals themselves. Thus, if a given terminal requires that each of its received packets contain an integer number of vocoder parcels, the concentrator will compose the packets accordingly. If a terminal merely expects a serial bit stream, either with or without encrypted portions, the concentrator may simply accumulate arbitrarily long segments of data for that terminal, and transmit them as necessary, without regard to parcel boundaries or other data details.

A somewhat more complex protocol is required for communication in the terminal-to-concentrator direction due to the shared nature of the channel. Two possibilities for this protocol have been considered; namely, an ALOHA/ETHERNET-type of contention mechanism and a slotted TDMA system under the direct control of the concentrator. Although a clear choice between these systems has not yet (and may never) emerge, several issues and potential trade-offs have been identified:

(1) An ALOHA-type scheme allows the terminals to transmit without regard to system time frame constraints. This permits each terminal to accumulate a parcel or a series of parcels that can then be forwarded as a single packet. Although this can usually be accomplished in TDMA schemes as well, timing requirements might result in more cumbersome protocols or in lowered transmission efficiencies. Although a Local Voice Network packet that contains an integer number of parcels might be retransmitted intact by the concentrator, it is not clear that this represents the only efficient way of handling parcel-oriented speech data. Another option might allow for arbitrary packetization of the data by the voice terminal, provided an easily identifiable pattern was included at the parcel boundaries. The concentrator could accumulate data from each terminal in a FIFO buffer and then determine the parcel boundaries by detecting that pattern. The latter need only be done for initial acquisition since presumably the concentrator will know parcel sizes for all the active voice terminals. Both the ALOHA and TDMA channel-sharing protocols appear to be equally attractive given this type of parcel identification scheme; however, the widespread use of variable parcel size algorithms might tend to favor the ALOHA method.

(2) The use of carrier sensing and collision detection is critical in the ALOHA/ETHERNET method. If separate cables are used for the two Local Voice Network channels, then the normal receiving hardware will be unavailable for these functions, and a separate hardware subsystem will be required for the carrier sensing and collision-detection operations. The same argument applies if two frequency bands are used on the same cable. The use of a time-shared strategy in which the cable is devoted to the concentrator-to-terminal broadcast function during one epoch, and used as an ALOHA channel for terminal-to-concentrator connectivity during the next, might allow for more efficient utilization of receiver equipment.

(3) A major advantage of a TDMA-based channel-sharing strategy is that one can avoid the possibility of collision completely. In this approach, the concentrator schedules the transmissions of the various terminals and communicates the required control information to the terminals via the broadcast channel. Burst transmission assignments can either be sent as separate broadcast messages or included as additional overhead in the normal concentrator-to-terminal data transmissions. A detailed example of a system design along these lines is presented in Sec. I-D-4-b.

(4) A major unknown for the ALOHA/ETHERNET solution is the statistical behavior of the system in a speech environment. In particular, the periodic nature of the voice sources might create collision or lockup difficulties that have not been experienced with Poisson-distributed data sources in existing ETHERNET systems.

b. Design Example

This section describes a Local Voice Network design based on the TDMA channel-sharing strategy. The design is offered primarily as a vehicle for identifying several important functions that have to be accommodated by the concentrator/access area/voice terminal system, as well as some hardware and software issues relating to voice terminals and speech concentrators in general.

(1) Transmission Medium

The major technical considerations in the design of a cable transmission system for the Local Voice Network relate to the cables used and the limitations imposed by modem designs and configurations. In the case of the ETHERNET, the medium used by both Farber at the University of California and Greenblatt at the M.I.T. Artificial Intelligence (AI) Laboratory was a standard low-loss 75-ohm coaxial cable available from the CATV community. One feature of this cable is the availability of cable taps known as Jerrold Taps for tapping into the cable at any point and introducing a transceiver at the interface to a terminal. This feature is especially attractive in that it facilitates complete terminal mobility such that terminals can be connected to, or removed from, the Local Voice Net with impunity. Although the concept sounds ideal, conversations with Metcalfe at PARC and Tom Knight at the M.I.T. AI Laboratory indicate that limitations are imposed by the non-ideal nature of the match between the cable and the transceiver. This results in a bound on the number of taps that can actually be supported by the system. While Metcalfe employed the Jerrold Taps, Knight resorted to separating the cable and affixing connectors to the ends in order to accommodate a transceiver. This was done as a result of some concern by the CATV community about the reliability of the taps. It is not clear whether the use of Knight's method would seriously compromise terminal mobility within an access area. Our current feeling is that the Jerrold Taps should probably be avoided.

System bandwidth is an important issue in the context of the number of voice conversations that can be supported, and in its effect on Local Voice Net protocol design. If local bandwidth were cheap and easily available, one might be able to exploit it in return for simpler voice terminal modem hardware. CHAOSNET bandwidth is 8 Mbps, but this is in an experimental system that presently links only three hosts over a maximum cable span of 1000 ft. ETHERNET experience indicates that bandwidths on the order of 2 to 3 Mbps can be safely realized in a system servicing virtually hundreds of voice terminals (not simultaneously off-hook). The ETHERNET concept is based on carrier sensing and collision detection. In this example, we are considering a TDMA alternative for the shared Local Voice Net channel, and can avoid potential collisions through the use of a concentrator-based scheduling mechanism. This leads to the possibility of dispensing with the carrier, given that there is no need for sensing it. The use of direct digital video signaling should result in less complicated terminals since the carrier generation and detection functions are eliminated. A potential problem area however, is that cable bandwidths may be less for video signaling than for carrier transmissions. The choice of two physically separate cables rather than a time- or frequency-shared single line should offer some relief in that regard.

(2) Communication Protocol

A slotted TDMA format is suggested for communication over both the terminal-to-concentrator and concentrator-to-terminal channels (Figs. 6 and 7). Time is divided into a

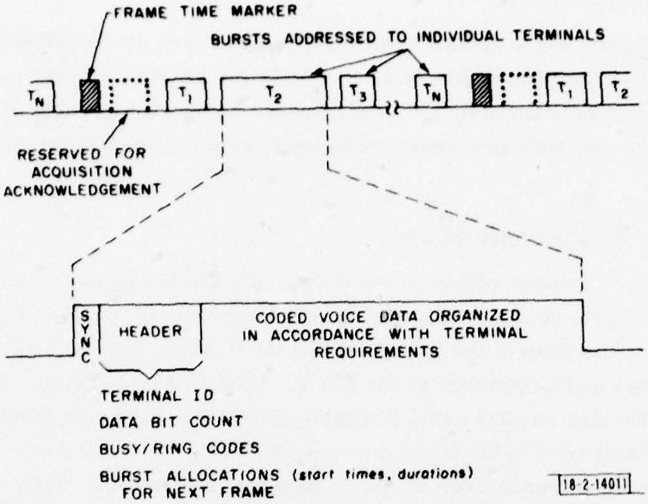

Fig. 6. Concentrator-to-terminal broadcast protocol.

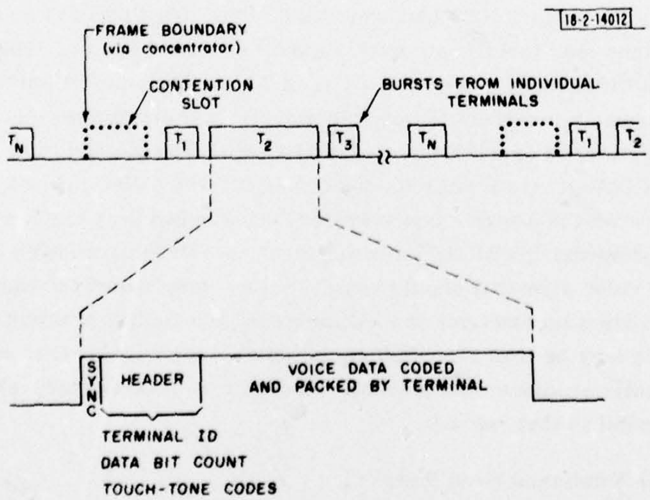

Fig. 7. Terminal-to-concentrator TDMA channel protocol.

sequence of frames of equal duration on both links. Frame boundaries are defined by marker patterns that are continually broadcast from the concentrator and recognized by all the speech terminals. Between successive frame markers, the concentrator transmits a sequence of individual bursts or packets addressed to specific voice terminals. Each terminal is aware, via a mechanism to be described below, of where in the frame its own data will be located. Although this may not be an overly important feature, it reduces receiver false alarms by allowing them to restrict the search for their data to a relatively small time window. The concentrator transmits one burst to every active voice terminal in each frame.

A similar strategy is used for terminal-to-concentrator transactions, except that a portion of the frame is reserved for contention signaling by terminals desiring to gain initial access to the system. In this scheme, a terminal that has heretofore been inactive begins by listening to concentrator broadcasts and locating the frame boundaries via the marker pattern. It then transmits its own identification code in the contention portion of the terminal-to-concentrator frame. Assuming that it was the only terminal to have done so in that frame, its code will be recognized by the concentrator, which in turn will respond by addressing a message to that terminal and sending it on the broadcast channel. This message will contain burst allocation information for that terminal to use for both listening and transmitting. All future transactions between that terminal and the concentrator will be conducted using those burst slots, thereby freeing the contention channel for use by other newly awakened voice terminals. A reasonable requirement might be that these "acquisition acknowledgment" messages (Fig. 6) be transmitted by the concentrator in a predetermined portion of the broadcast channel frame. This offers false-alarm protection during the acquisition phase, and allows for the use of relatively short synchronization headers. In the event that more than one terminal transmits in the contention slot simultaneously, the concentrator will be unable to acknowledge any of them. Each terminal might retransmit its ID after a random waiting time following a given time-out interval.

In general, the receiving and transmitting burst positions for a given terminal need not be the same. In fact, under adaptive variable-rate voice strategies the transmitting and receiving bandwidths (e.g., burst widths) of a terminal will often differ. In addition, as old conversations are terminated and new ones are initiated, some relocation of the burst assignments might be appropriate. The normal concentrator-to-terminal data protocol also includes burst allocation information, thereby allowing the concentrator to dynamically modify burst positions in both channels as a function of time. We note that the role of the terminal is to extract burst allocation information from its received data stream, and then to simply count time from each frame boundary until its assigned receiving and transmitting slots appear. The hardware required for implementing these functions should result in fairly small and inexpensive terminal modem designs. The more complicated scheduling functions have been relegated to the concentrator, where they need be implemented only once and shared among the many voice terminals.

The following comments and observations can be made with respect to the above-described system:

(a) The use of a recurring frame structure guarantees a regular flow of data to and from the speech terminals during periods of speech activity. This reduces the amount of buffer memory that might be required at the terminals in order to accumulate packets for transmission, or to store them upon receipt.

(b) The explicit inclusion of touch-tone codes in the terminal-to-concentrator protocol permits the use of the keyboard for signaling during the course of a conversation. This is an important requirement for structured voice conferencing applications.

(c) Dial-up and call-termination logic resides in the concentrator. The communication protocols merely relay touch-tone keyboard inputs to the concentrator, and call status codes to the terminals. This minimizes terminal complexity and avoids the unnecessary duplication of logically complicated but infrequently used functions in the system.

(d) Bit counts are included in burst headers to account for possible variations in the actual amount of data that is sent in successive bursts to or from the same terminal. This condition can be expected when independent timing considerations coexist in the same system. In this case, the Local Voice Net frame rate is independent of voice terminal bit rates or vocoder parcel rates.

(e) Terminal IDs are included in the headers as an aid in recovering from possible system problems. Their presence allows the concentrator to verify that terminals are performing according to instructions, or to identify those that are not.

(f) Local Voice Net frame durations of between 20 to 50 msec seem reasonable.

(g) Although propagation delay differences between various terminals and the concentrator are expected to be small for the access areas in the wideband experiment, the efficient use of the shared TDMA channel might be affected by this phenomenon in larger systems. We observe that close "packing" of bursts from different terminals can be organized by the concentrator by making use of the system's dynamic allocation feature. For example, if a burst from a given terminal is arriving too late or too early, the concentrator can suitably modify the "start of burst" parameter in its next transmission to that terminal.

(h) A logical equivalent of the Local Voice Net can be constructed by running separate cables from each terminal to an ohmic junction point at the concentrator. This has the appearance of the star geometry, but it does not require a separate I/O interface for every line. Although it might involve higher wiring costs, this configuration affords the concentrator the opportunity to selectively disconnect a terminal that might be misbehaving due to hardware failure.

E. Voice-Terminal Definition

In this section, we present a voice-terminal structure that satisfies the design criteria outlined in Sec. C. The structure is canonical in that it can represent a variety of terminals designed for use with different access area/concentrator systems via the appropriate definition of

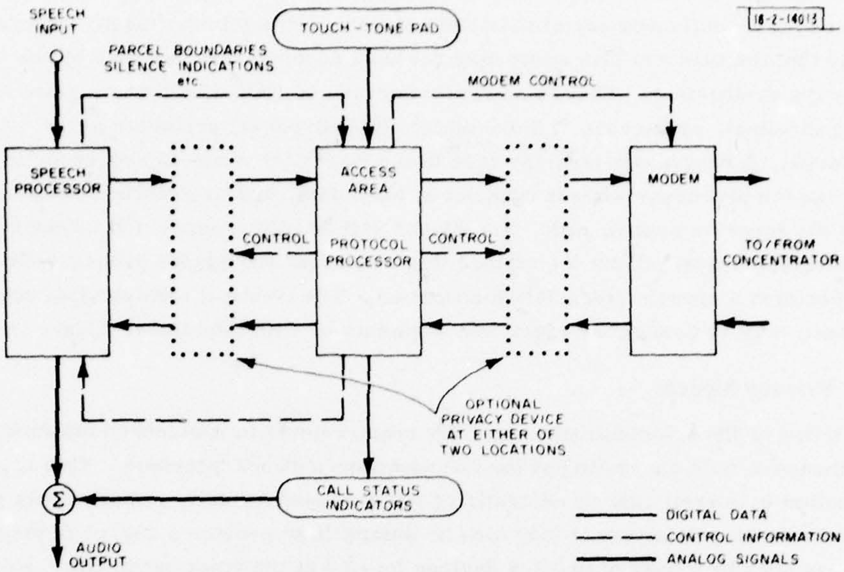

Fig. 8. Canonic voice terminal structure.

its functional blocks. Specific requirements are discussed in the context of the above-described Local Voice Net design.

The block diagram of Fig. 8 contains four major elements. Although detailed design studies have not yet been conducted for these various subsystems, we discuss below several considerations that can potentially impact their overall size, cost, and complexity.

1. Speech Processor

This is probably the single largest element of the voice terminal, at least in the case of narrowband speech. It might account for 50 percent or more of the total terminal hardware. We anticipate that the most convenient format for the digitized voice I/O will be bit serial, with separate physical ports for the input and output streams. Both streams are continually clocked at a constant rate determined by an internal speech processor clock. The latter may be used by other voice terminal subsystems (e.g., the privacy device) if necessary.

Two considerations point to the use of serial rather than parallel (word transfers) speech processor I/O. First, we expect that in the not too distant future, speech processors will reside on several custom-made LSI chips. A major limitation of LSI systems is in the number of leads that can be comfortably provided for external connection. Serial I/O reduces this requirement to a manageable level. Second, variable parcel size systems and future variable rate or embedded coding methods lack the uniformity of data structure that might benefit from the use of fixed word length parallel I/O.

Note that parcel boundary markers and silence indications are separately provided, eliminating the need for bit-by-bit searching of the data by the protocol processor. This also permits the serial data to be scrambled without denying the protocol processor the opportunity to perform TASI-like functions or parcel-oriented data formatting on the scrambled information.

In the receiving direction, the speech processor functions as a conventional vocoder synthesizer, accepting a continuous serial input data stream. Provision is made, however, for the possibility that the protocol processor may not have received any new data by the time they are needed by the synthesizer. This could result because of TASI-like transmission activity at the sending terminal, or because of delay effects or lost-packet problems in the external wideband network. A reasonable strategy here would be for the protocol processor to forward "junk" to the speech processor when it runs out of valid data, and to simultaneously indicate "silence" via the separate control path. The "junk" will be unscrambled and turned into yet another meaningless sequence, but the silence flag will cause the speech processor to ignore the data and perform a speech interpolation operation. The identical interpolation procedure will work equally well in dealing with lost data segments or with intentional silence intervals.

2. *Privacy Module*

A large share of the communication security requirements in a wideband network could be provided by means of bulk encryption at the concentrator/network interface. This approach offers a reduction in overall cost by centralizing the stringent security requirements and allowing simpler terminals. However, it may also be desirable to provide a degree of privacy within the local access area by means of privacy devices located at the voice terminals. More expensive terminals for true end-to-end encryption could be provided to the few individuals or locations that actually need them.

Two important issues arise in considering the inclusion of privacy modules at the terminals: key distribution and synchronization. The key-distribution problem relates to the fact that private communications require that the conversing terminals have compatible keys. For communication between different access areas, the assignment and distribution of keys to the terminals would have to be controlled by their respective concentrators. The transmission of the keys from concentrators to terminals in a private manner implies special requirements on the protocol processors at each terminal.

If end-to-end privacy is to be maintained between voice terminals communicating over a packet network (or any network), then provision must be made for acquisition and maintenance of privacy-device synchronization between the two terminals, in addition to the usual parcel synchronization required for speech communications. The most satisfactory arrangement would be to handle these two types of synchronization in an integrated fashion, and in such a way that synchronization is maintained despite packet losses in the network.

One approach to dealing with these issues is to utilize, more or less directly, the BCR (black-crypto-red) technology currently under development by BBN and others. This technology provides privacy (including the incorporation of key distribution and privacy device synchronization) between host computers in a packet network in a manner which is transparent to the host computers. Direct application of this approach to end-to-end speech privacy requires that the speech terminals perform all network host functions as well as the usual speech functions. For example, conferencing and dial-up protocols would have to be accommodated in the speech terminal. The concentrator would simply serve as a gateway, and forward terminal packets to the wideband network. Referring to Fig. 8, the protocol processor would be carrying out the host function and the privacy device would take the form of a BCR processor located to the right of the protocol processor. This approach is at an extreme in terms of separation of functions between terminal and concentrator, representing a maximum in cost and complexity of the terminal. It also leads to increased cost in the area of communications overhead, since a full network or internet header would have to be included with every packet leaving the protocol processor, and the data portion of the packet would have to include leader and padding bits to provide for independent privacy device synchronization for each transmitted packet. However, the approach does represent an existing solution to the privacy problem and a clean separation of the speech and networking functions of the terminal from the privacy functions.

We have focused our attention in this report on functional partitionings that offer greater simplicity in the voice terminals than appears to be possible using a BCR-based end-to-end privacy strategy. In this regard, it is worthwhile to consider other approaches to end-to-end privacy in which adherence to strict privacy requirements may be less stringent, but which offer the potential of less complex, cheaper terminals. One such scheme would be to scramble the voice bit stream with a bit-oriented privacy device placed between the speech processor and the protocol processor (see Fig. 8). Information relevant to speech packetization such as parcel boundaries and silence indications would not be scrambled, and could be passed on to the concentrator in the clear if necessary. In addition, touch-tone signaling information could pass to the concentrator in the clear so the network dial-up and conferencing protocols could be implemented in the concentrator. Of course, this information could undergo backbone encryption at the network side of the concentrator. Synchronization of the privacy devices could be accomplished by having the protocol processor time stamp its transmitted parcels in such a way that a receiving processor could determine whether any speech parcels were lost in the network, and advance its privacy device by enough steps to stay in synchronization. A detailed review of some of the properties of encoded speech streams and several possibilities for achieving joint vocoder and privacy synchronization are presented in the Appendix. An approach in which some of the network host functions are physically separated from the voice terminal as described above, must be coupled with a method for distributing keys to the terminals, via the concentrator, in a private manner. The implications of this requirement on the terminals and concentrator remain a subject for further investigation.

3. Protocol Processor

The function of this subsystem is basically to control and format the flow of data into and out of the terminal. Using the data formats of Figs. 6 and 7 as examples, the protocol processor composes header information, appends it to appropriately chosen segments of the voice data stream, and forwards the augmented information to the modem for transmission to the concentrator. In the receiving direction, the protocol processor separates voice data from header and synchronization bits, and creates a continuous serial data stream for the speech processor. In the event of missing data due to TASI or lost segments, the protocol processor provides a "silence" indication to the vocoder synthesizer.

An additional function of the protocol processor is to control the timing and the operation of the modem. In the above-described TDMA-based Local Voice Net example, the protocol processor would be responsible for converting burst allocation parameters received as part of the concentrator-to-terminal protocol, into appropriate modem control signals.

A natural candidate for implementing the protocol processor is a microprocessor system or a one- or two-chip microcomputer. However, these typically deal with parallel word operations, and might be poorly matched to the serial bit streams flowing to and from the speech processing portions. An interesting possibility might be to architect a combined serial/parallel structure in which the through-flowing serial data remain in serial form, while the header and protocol bits are formulated and manipulated in a conventional microprocessor. This would relieve the microprocessor of handling the relatively high voice data throughput rate, and could result in some hardware economy. A necessary ingredient in this design would be a structure for inserting newly formed headers and synchronization patterns into a through-flowing bit stream and vice versa.

A subject that requires additional study relates to the possibility of dealing with network delay dispersion and/or packet order inversions in the terminal. This function can potentially be accommodated in either the terminal or the concentrator, and the choice depends largely upon issues of system economy and flexibility. We note the following:

(a) This function is an ongoing one for all active voice connections. Centralized implementation does not therefore offer the same economic advantages as in the case of dial-up logic or conferencing protocols, which are infrequently used by individual terminals. However, the delay compensation and packet reordering functions are needed only for those terminals that are off-hook, and these will generally constitute a small fraction of the total number that are deployed in a given access area. The economy of centralized implementation might therefore still be significant.

(b) Packet order inversion and delay compensation algorithms depend upon time stamps for their operation. If these functions are performed at the terminal level, then a terminal-generated time stamp has to be included in the terminal transmission format. This represents an additional terminal function, but may not be unreasonable in that a natural time base for the time stamps is the parcel unit, which is produced by the terminal.

(c) Although system economy and flexibility will ultimately determine where given functions should be performed, we observe that delay dispersion and packet order inversions are network-induced effects rather than specific speech-related issues. Indeed, two speech terminals in the same access area should be able to communicate with each other without the need for packet reconstitution algorithms, even if their transmission formats are packet oriented. This follows from the basic requirement that access areas provide simple connectivity between terminals and concentrators, without introducing deleterious side effects of their own.

4. Modem

The modem is responsible for converting the digital output of the protocol processor into a form suitable for transmission to the concentrator, and for converting received concentrator signals into digital form for the protocol processor. For our TDMA-based Local Voice Net example, the modem might be little more than a pair of serial shift registers that can be loaded at one rate and unloaded at another, with timing controlled by the protocol processor. The receiving portion might contain a special-purpose synchronization-pattern recognition filter, in order to gain rapid acquisition of an incoming burst. For ETHERNET-like access areas, the modem will require carrier generation and detection hardware and collision-sensing and recovery capability in addition to the burst-forming circuitry.

F. Concentrator Functions

The role of the speech concentrator is to act as an interface between a local community of voice terminals and a wideband digital integrated network. Since a number of speech concentrators will be connected to different nodes in the wideband network, it seems reasonable to require that those portions of the concentrator designs that deal with network protocols be more-or-less identical. On the other hand, access-area requirements might dictate different designs for some installations than for others, and some concentrators might therefore be configured quite differently from others.

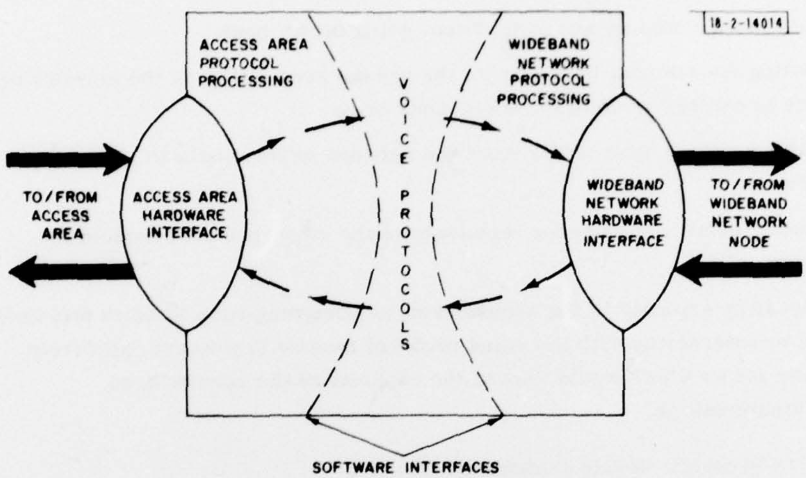

Fig. 9. Speech concentrator functions.

Referring to Fig. 9, we have partitioned the concentrator into five major functional areas. Two of these deal with hardware interfacing – to an access area on the one hand, and to a network node on the other. Depending on the details of the access-area protocols, the hardware interface between the access area and the concentrator might be designed to relieve some of the computational burden that would otherwise fall on the concentrator. For example, in the case of the Local Voice Net, the interface might include filters that are matched to synchronization patterns, hard-wired registers for unpacking header data, etc. This approach assumes a certain amount of stability in the access-area protocols, since modifications are less easily accommodated than via software alone. However, the access area contains large numbers of terminals in which the protocols also exist, and for which stability has to be assumed in order to achieve low-cost designs with present-day device technology. There thus appears to be little advantage in restricting the access-area functions of the concentrator to software implementation alone. Similar arguments could be made for including some special-purpose hardware in the network interface hardware. In the case of the ARPA/DCA test-bed system however, this might limit the flexibility of the system for networking experiments.

With regard to software design and functional partitioning, several suggestions and examples can be found in the previously referenced NSC Note No. 123. The main point that we emphasize here is that it should be possible to design software interfaces between the access-area-specific protocols, the voice protocols, and the network-specific protocols such that, given those interfaces, the various protocols can be written independently of each other. This would allow for similar network protocol software in all the concentrators while accommodating a variety of access-area designs at different locations.

Listed below are several of the functions that would be required in a concentrator. The listed access-area-related functions are somewhat tailored to the TDMA example of Sec. D-4. However, the remaining functions of the concentrator are not related to a specific access-area structure.

1. Access-Area-Related Functions

(a) Monitoring terminal status to detect going on/off hook

(b) Routing speech data bursts from the access area either to the network or back to another terminal in the access area.

(c) Routing speech data bursts from the network to terminals in the access area.

(d) Routing control signals for terminals to the voice protocol module for action.

(e) Allocating capacity in the access area by assigning time slots to terminals and by interacting with the voice protocol module to prevent calls from being set up which would exceed the capacity of the access-area communications.

2. Voice Protocol Module Functions

(a) Setting up calls. This function would involve the following steps:

(1) Engaging in a dialog with the user. User key pushes would be sent via control signals from the terminal to the protocol module. Signals in the reverse direction would produce audible tones or lights to indicate the state of the call set up, i.e., dial tone, ringing, busy.

(2) Negotiating with the network protocol module and/or the access-area module to obtain the communication resources necessary to handle the call.

(3) If the call involves a remote concentrator, negotiating with the voice protocol module in that concentrator (which in turn negotiates with its access-area module) to determine whether resources are available at the destination and whether or not the called terminal is busy.

(4) If the call involves end-to-end privacy, arranging for negotiations between the involved terminals and a key distribution center to get a privacy key for the call.

(5) If the terminals are capable of operating at a variety of bit rates, negotiating with the terminals to select a rate acceptable to the terminals and compatible with available network capabilities.

(b) Taking down calls when either the local access area or the remote protocol module indicates that one or the other party has hung up. The relevant access area and network modules would be notified so as to free the previously committed resources.

(c) Supporting voice conferencing. There are many options for voice conferencing. Control may be centralized or distributed and may make use of speech activity detection or special control signals. In this study, we have not focused attention on one or another technique. Depending upon the technique used, the voice protocol module would take appropriate action.

(d) Accounting. In a network with paying customers, the voice protocol module would record the appropriate information so that the customers could be billed.

3. Network-Related Functions

(a) Monitoring network status and measuring (or estimating) available capacity so that network congestion can be avoided by denying call setups which would cause overloads or requiring new calls to use lower data rates.

(b) Packetizing (and depacketizing) speech data bursts in formats suitable for the wideband network. This process may involve aggregating bursts from many speakers into large network packets to gain network efficiency without the increased delay which would result from accumulating a large packet's worth of speech from a single speaker.

II. ADAPTIVE VARIABLE-RATE PACKET SPEECH NETWORKING

A. Introduction

A rate-adaptive packet speech network strategy based on an embedded speech coding technique and a variable-rate communications protocol was described in a previous report.* In this report, initial work on a simulation intended to investigate the behavior of such a system from a networking viewpoint is described. Parallel efforts in the ARPA-sponsored Packet Speech Program (reported in the current Packet Speech SATS) have led to a very promising variable-rate speech encoder which is compatible with the embedded coding approach assumed here.

*Information Processing Techniques Program Semiannual Technical Summary, Volume II: Communications-Adaptive Internetting, Lincoln Laboratory, M.I.T. (31 March 1977), DDC AD-A044071.

For the purposes of the network experiment, it has been assumed that each voice user has a vocoder which generates bits at a total rate of 16.8 kbps, and that subsets of these bits support speech synthesis at seven different rates ranging from 2.4 to 16.8 kbps in equal increments. A priority-oriented packetization scheme with seven priority levels corresponding to the seven bit rates is assumed. For example, if only the highest priority (priority 7) packets are reaching the receiver, then speech synthesis at 2.4 kbps is supported; if packets of priority classes 3 to 7 are being received, 12-kbps speech synthesis is possible. Network nodes allocate their transmission capability based on these priorities. When overload conditions exist and queues begin to build up, low-priority packets are discarded until the overload is relieved. The quality of the synthesized speech is determined by the lowest priority packets that reach the receiving terminal. Feedback schemes can be implemented for end-to-end flow control, where the receiving terminal sends information to the transmitting terminal concerning the packet priorities currently being received. Then the transmitting terminal can lower its transmitting rate to avoid loading intermediate nodes with low-priority packets which are not reaching the receiver.

B. Network Configuration

The network configuration selected for an initial simulation is one which allows reasonably simple implementation yet is complex enough to allow for experimentation with feedback schemes supporting end-to-end flow control. As shown in Fig. 10, the modeled network consists of a central node through which pass 16 paths connecting 4 nodes on either side of the central node.

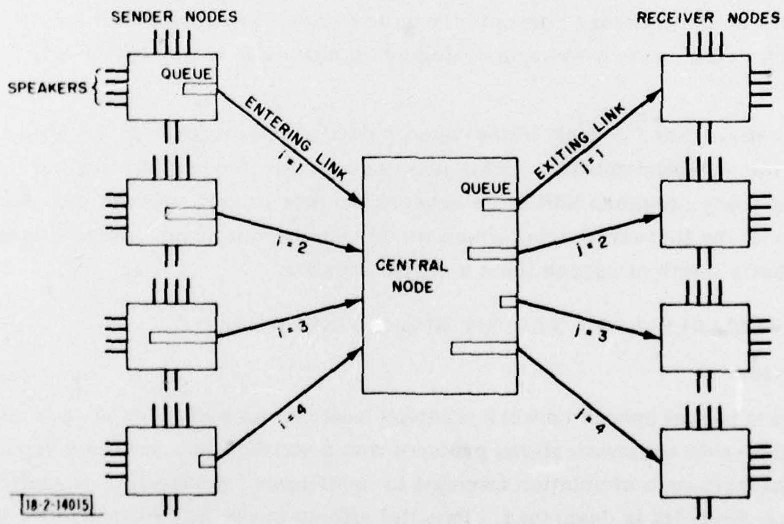

Fig. 10. Simulated network used to test embedded coding/bit-stripping schemes.

The central node is assumed to maintain independent queues for each outgoing path. It is assumed that packet voice terminals are connected to all nodes except the central node. The traffic is specified by a matrix which indicates how many voice terminals at node i on the left of the center are in conversation with terminals at node j on the right. A fixed traffic matrix is assumed, but fluctuations in packet production due to individual talkers oscillating between talkspurt and silence are included in the simulation. A statistical talker activity model* is used to control the talkspurt/silence behavior of all talkers. Although all conversations would actually be two way, the simulation deals only with the traffic proceeding from left to right. Thus the nodes on the left are viewed as "sender" nodes, and the nodes on the right are "receiver" nodes.

The simulation proceeds on an event-by-event basis, where an event consists of the initiation or termination of a talkspurt by one of the talkers. The system state, which is updated after every event, is specified by the following variables:

(1) The number of talkers in talkspurt on path (i, j) connecting the i^{th} sender node to the j^{th} receiver node;

(2) The number of packets in each of the queues at each sending node and in each of the four queues at the central node;

(3) The lowest priority packet currently being transmitted on each link in the system; for example, if link 2 leaving the central node is currently supporting priority 3, then packets of priority 1 and 2 entering the central node and targeted for receiver node 2 will be discarded by the central nodal processor;

(4) The lowest priority packet currently being delivered by each terminal to its sender node (this can be different from priority 1 in the cases where end-to-end flow control is employed).

When a talker on the path (i, j) enters or leaves talkspurt, the i^{th} sender queue and the j^{th} central queue are updated. Estimates of the sizes of these two queues are then projected 1/2 sec into the future, based on the current bit rate. If the projected size of either queue exceeds a threshold, the lowest priority packet accepted by the corresponding link is increased by 1. If the projection indicates that the queue is emptying at such a rate that the link could support a higher transmission rate, the lowest priority packet accepted is decreased to the next lower level.

The system is initialized with all terminals sending at maximum rate, and all links accepting even the lowest priority packets. The available data rate is set to be grossly inadequate to support all terminals at full rate. As time passes, the links strip off lower priority bits in response to excessive queue sizes. After a few tenths of a second, an approximate steady state condition is reached, and rates are adjusted only infrequently thereafter. Queues remain small such that delays at each link are never greater than 0.1 sec.

C. Feedback Schemes

There are currently three variations of the main system which are being investigated and compared as to response to identical initial conditions. In the first and simplest system (no feedback), sending terminals always send at maximum rate, and receivers receive at the maximum rate possible, given the load on each link in the path. In the second system (continual probing), the receivers communicate back (via short control packets) to the senders the average

*Information Processing Techniques Program Semiannual Technical Summary, Volume II: Communications-Adaptive Internetting, Lincoln Laboratory, M.I.T. (31 March 1977), DDC AD-A044071.

rate received over the past fixed small time interval. The senders respond to such updates by setting their sending rate to be the next higher rate above the average received rate. Delays are introduced for the feedback messages, equal to twice the sum of queue delays at each of the two links in the path (assuming equivalent link loading on the return path). The third and most complicated strategy (periodic probing), consists of the sender sending at the maximum rate received by the receiver, again communicated back after a time delay dependent upon queues. Such a system would never respond to an easing up of the network load. Thus an up-probing feature is added such that if after a certain elapsed time interval, the maximum received rate has remained equal to the sending rate, then the sender up-probes by increasing his rate by one level. The times for this system have been set such that the receivers update every 0.4 sec, and senders probe after a steady period of 1.2 sec. For both of the feedback systems, it is assumed that all voice terminals on a given path operate in unison.

The third system has the disadvantage that it is not as responsive to a decrease in the load as are either of the other two systems. However, it is capable of sustaining a higher average rate because, in the ideal, sending rates and receiving rates will exactly match and no bits will be discarded in the network. Hence, early nodes in the network path are not as heavily loaded, and the possibility exists that delays and/or received rate at other nodes can be improved.

D. *Experiments*

A few experiments have been run to investigate the behavior of the system under the different feedback/flow control schemes. The indications are that the feedback schemes are effective in adjusting terminal transmission rates to account for large imbalance in traffic flow, but have little effect when the traffic matrix is more uniform.

In a first run, the traffic matrix was arranged such that all links entering the central node were equally loaded, whereas the load on the exiting links increased linearly from 36 speakers on the uppermost link to 72 speakers on the lowermost. The hope was that in the case of feedback the receivers on more heavily loaded exiting links would communicate back to senders on all four entering links that they were receiving at a low rate. The corresponding senders would then reduce their rate accordingly, and hence allow other users of the shared sender link to up their rate. The result would be improved overall quality of speech received across the less heavily loaded receiver links.

The results obtained were that neither of the feedback mechanisms gave significantly improved rates over the system with no feedback. By the end of 2.5 sec, the system with no feedback was actually producing higher average rates received than the system with periodic probing. This effect was due to an overshoot phenomenon in the delayed feedback. After 3.5 sec, the periodic probe system had recovered, and it sustained a slight edge over the non-feedback system for the remainder of the run. The improvement resulting from the system with continual probing was so insignificant as to be completely discounted.

The three systems were then compared when run with a much greater imbalance in the data flow. For this run, the traffic matrix was arranged such that 10 speakers were talking from each sender node to each of the 3 upper receiver nodes, and 70 speakers were talking from each sender node to the lowermost receiver. For this case, both of the feedback systems allowed for dramatically improved quality of speech arriving to each of the three upper receivers, over that allowed by the system with no feedback. The data rate was set so that only the highest priority packets could get through the last receiver link. The feedback mechanisms thus resulted in a drastic reduction in the number of bits sent over the links entering the central nodes. This allowed these links to operate at a higher overall priority level. The results at steady state were that, in the case of no feedback, sender links rejected all packets of priority lower than 5. With constant probing, sender links could accept packets down to and including priority 3, and sometimes 2. With periodic probing, the gains were even better, with sender links able to accept even priority 1 packets, most of the time. However, it took 8 sec for the periodic probing system to reach a reasonably steady state condition.

Future plans call for further investigation of a variety of feedback schemes under different network load conditions, along with the development of performance measures to evaluate the various schemes. A set of display routines designed to allow for comparisons among various alternatives will be designed. In addition, a coupling of the network simulations to a variable-rate embedded-coding vocoder (described in current Packet Speech SATS) is planned. The vocoder rate would be varied in real time as if it were one of the adaptive voice terminals in the network, and effects on perceived quality of frequent changes in rate could be evaluated.

APPENDIX

SYNCHRONIZATION ISSUES IN PACKET SPEECH COMMUNICATION

I. INTRODUCTION

In order to carry out digital speech communication, it is of course necessary that the analog speech be analyzed and encoded into digital form at the transmitting terminal and decoded and synthesized into analog form at the receiving terminal. It is also generally required that a terminal-to-terminal synchronization with respect to the structured format of the digitized speech data be established and maintained. If end-to-end privacy between terminals is to be accommodated, an additional synchronization requirement relating to the privacy devices is introduced. These synchronization requirements are relevant to circuit-switched as well as packet-switched environments. However, the special nature of packet speech communications makes the synchronization issue somewhat different in the packet environment. In addition, it is convenient and desirable in packet speech systems to save on channel utilization by not transmitting packets during silence intervals, and the accommodation of this feature tends to become coupled with the synchronization problem. The purpose of this appendix is to discuss the issues of speech-stream synchronization, privacy-device synchronization, and silence detection in a packet network and to indicate methods by which the terminal/concentrator communications format can support these functions. The discussion begins with a review of the structural properties and implied synchronization requirements of encoded speech streams. Then, some general methods for applying privacy transformations to digital bit streams are reviewed. Finally, a few example configurations of speech encoders and privacy devices are described, and terminal/concentrator communication formats suitable for each configuration are set forth.

II. PROPERTIES OF ENCODED SPEECH STREAMS

The serial bit stream produced by a speech encoder generally has a structured format so that synchronization with respect to this format must be established and maintained between encoder and decoder. Typically this format is periodic in that the encoder produces fixed-size blocks of bits, called parcels, at a uniform rate. The possible range of parcel sizes is rather large, as indicated by the following examples:

(a) 16-kbps APC – 320-bit parcel every 20 msec,

(b) 2.4-kbps LPC – 49-bit parcel every 20 msec,

(c) 64-kbps PCM – 8-bit "parcel" (speech sample) every 125 μsec.

An example of a technique for acquiring and maintaining parcel synchronization (when two terminals communicate only via serial bit streams) applicable to low-rate vocoders, is to transmit a known bit pattern (corresponding to an illegal pitch value) in place of the pitch word during unvoiced utterances. This pattern can be searched for continuously in the serial stream and synchronization declared when the known pattern has been found at the same location in a sufficient number of adjacent parcels. This technique is sluggish, but it is applicable in circuit-switched environments where synchronization loss is a very rare occurrence. If loss of bit integrity is a frequent phenomenon (e.g., in packet networks), it might be desirable to speed up the acquisition of parcel synchronization (at a cost in overhead) by adding a fixed number of synchronization bits to each parcel.

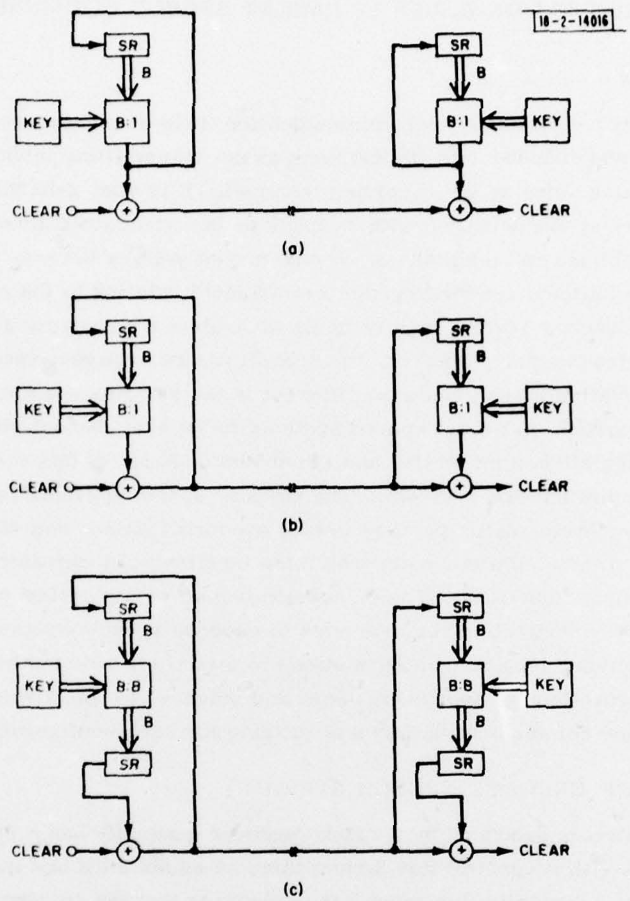

Fig. A-1. Privacy-device models. (a) Bit-by-bit, data-independent technique; (b) bit-by-bit, data-dependent technique; (c) block-oriented, data-dependent technique.

In packet-switched systems, the maintenance of parcel synchronization can be assured if an integer number of parcels is included in every packet. This seems to be a reasonable approach to take if the voice terminal transmits to the concentrator in packetized form and the packetizer has access to parcel boundaries. If parcel synchronization is to be maintained in this way, then it may be necessary for the concentrator to fragment speech terminal packets for transmission on the network. For example, networks with small, fixed packet sizes carrying on the order of 100 information bits* have been proposed to reduce delay and overhead for packet voice communication. If parcel size is 300 bits and speech terminal packets contain whole parcels, then network packets must be fragments of terminal packets. Since the packets exchanged between terminal and concentrator are of private concern to the local access area, fragmenting and reconstruction of these packets can be carried out locally by the concentrator, and are not of concern to the external network.

Speech algorithms have been devised in which the encoded speech bit stream does not have a periodic structure, as in the case of vocoders with variable parcel size. Such vocoders depend to a great extent on the "free" parcel synchronization provided by packet boundaries. If this synchronization were not provided by the network, then adding the necessary synchronization information to the serial stream would cut down the bit-rate advantage of variable parcel size vocoders with respect to fixed parcel size vocoders.

Finally, there has recently been much interest in multi-rate speech encoders of the "embedded coding" variety, where different size subsets of the bits produced in every parcel interval can be used to support speech synthesis at a variety of rates. In a packet environment, an appropriate strategy would be for the terminal to organize these different sets of bits into separate priority-ordered packets for transmission to the concentrator. Based on observed network performance, the concentrator can decide on the quality of service achievable for each terminal and forward packets of the appropriate priority levels to the network.

III. PRIVACY-DEVICE MODELS

Some general privacy-device models are illustrated in Figs. A-1(a-c). Figure A-1(a) depicts a bit-by-bit, data-independent scheme where a pseudorandom sequence produced by a shift-register feedback arrangement is added (modulo 2) to the data stream at transmitter and receiver. Assuming that the B-bit to 1-bit transformation logic (which is dependent on a key) is the same at transmitter and receiver, and that both B-bit shift registers are initially loaded with the same contents, the received bit stream will exactly match the input bit stream. A single-bit error on the channel will result in a single-bit error at the output, but a loss of bit count (via packet loss or any other mechanism) on the channel will result in garbled output until the problem is detected and some mechanism for resynchronization can be initiated.

In Fig. A-1(b), the modified data are fed into the shift register both at transmitter and receiver. A single-bit error on the channel can cause up to B bit errors in the output data. However, the system will automatically resynchronize within B bits after a loss of bit integrity on the channel, since transmitter and receiver shift register contents will always match after B bits have been communicated correctly.

* J. W. Forgie and A. G. Nemeth, "An Efficient Packetized Voice/Data Network using Statistical Flow Control," Proc. IEEE International Communications Conference, ICC 77, Vol. 3, pp. 44-48 (June 1977).

Figure A-1(c) illustrates another self-synchronizing scheme, where a block-oriented transformation unit (labeled B:B) is employed. This unit accepts a block of B input bits and delivers B output bits, with the exact transformation being dependent on a key which must be available at transmitter and receiver. An example of such a device is the Digital Encryption Standard of the National Bureau of Standards, which operates on a 64-bit block and uses a 56-bit key. The top shift registers in Fig. A-1(c) are filled in bit serial fashion and fed in parallel once every B-bit interval through the block transformations. This produces B new bits for the lower shift registers, which are clocked out serially to combine with the input data bits. If bit count is lost in this system, it is necessary to reestablish block synchronization. Then transmission of one B-bit block through the system will put the transmitter and receiver shift registers in the same state so that subsequent data will be received correctly.

The bit-by-bit approach illustrated in Fig. A-1(a) has been the one most often used for voice communications in a circuit-switched network. Block-oriented approaches have been applied more frequently for block-oriented data communication. Although the continuous bit-serial nature of the speech processor I/O appears to be well matched to bit-by-bit scrambling techniques, the transmission formats in packetized voice networks lend themselves to block-oriented methods. Both approaches thus ought to be considered as possible alternatives for packetized speech.

IV. SAMPLE TERMINAL CONFIGURATION AND COMMUNICATION FORMATS

The communication format between terminal and concentrator should support parcel synchronization, privacy-device synchronization, and maintenance of correct silent durations in the face of possible lost packets. The preferred result is to achieve all these goals in a unified manner, but various approaches will be needed depending on such factors as the configuration of the terminal, the type (if any) of privacy device employed, and whether or not speech transmission is to cease during silent intervals. Three examples are given below to indicate the range of possibilities. The issues of whether packets are actually formed in the terminals or the concentrator has been left open except in the third example, where a packetized terminal is assumed. In the first two examples, it is assumed that if the terminal does not carry out a packetization function, it must exchange enough information with the concentrator (parcel boundaries, bit counts, silence indicators, etc.) to allow the concentrator to carry out a packetization function equivalent to that described.

A. Serial Stream Encoder, Parcel Boundaries Unavailable

Suppose we are presented with a voice terminal that produced a serial stream which is modified bit-by-bit (for privacy) in a manner similar to that indicated in Fig.A-1(a). The task is to transmit this stream over a packet network in such a way that packet loss will cause minimal degradation in the output speech. The packetizer has access only to the transformed serial stream and thus cannot determine parcel boundaries. Loss of a packet will cause loss of privacy-device synchronization for an indefinite period unless some action is taken. One possible aid would be to include information in the transmitted packet which would enable the receiving depacketizer to determine exactly how many data bits were lost if packets were missing. For example, transmitted speech packets could be required to contain a fixed number of information bits and could be augmented with sequence numbers. Then the depacketizer could transmit dummy bits if necessary to insure that bit count integrity is maintained between transmitting and receiving terminals. The dummy bits would cause packet-duration glitches in the output speech, but would prevent indefinite loss of privacy-device synchronization.

Note that this approach also insures that parcel synchronization, once established between the two terminals, will be maintained despite the loss of packets which do not contain integer numbers of parcels. Thus the approach is relevant even if privacy transformations are not carried out. It provides a general method for providing a packet interface to a serial-oriented vocoder without bit-by-bit processing of the serial stream to detect parcel boundaries.

B. Bit-by-bit Data Transformation, Parcel Boundaries Available

Suppose that a bit-by-bit, data-independent privacy transformation is used, but parcel boundary information is available in the clear along with the transformed data. Assume initially that silence detection is not used. One approach to maintaining both privacy-device and parcel synchronization (assuming fixed-size parcels) is to always pack an integer number of parcels into a packet and to indicate in each packet a sequence number (representing a time stamp in terms of inter-parcel intervals) of the first parcel transmitted. When packets are lost, the receiving privacy unit could be advanced by enough bits to account for the known number of missing parcels, and the receiving speech algorithm processor could be told (via side information transmitted in the clear) how many parcels were missing. An appropriate speech-algorithm-dependent strategy could be used for filling in these parcels in a manner which degrades the output speech as little as possible (see Sec. E-1). If speech activity detection is employed and parcels are not transmitted during silence, this same scheme will produce the required silence intervals. The privacy units at both ends of the channel could be made to continue clocking during silence, and time stamps would be placed on outgoing packets as if transmission were continuous. Both lost packets and silence intervals would result in an observation of missing parcels at the receiver, and in either case the time stamp could be used to keep privacy-device synchronization. Here the time stamps would also serve the role of maintaining correct silence duration intervals in the face of variable packet delay.

C. Block-oriented Transformation, Following Packetization

The model here is that the privacy transformation is carried out at the terminal on a packet-by-packet basis. This approach is supported by the BCR technology referred to in Section E-2, and some of its features and costs were discussed in that section. A B-bit block [see Fig.A-1(c)] is placed in front of the block of data bits to be transmitted in each packet and padding is added to insure that an integer number of B-bit blocks is sent. Time stamps or other control information could be included in the packet and scrambled along with the vocoder bits. Since packets are scrambled independently, privacy-device synchronization is transparent to speech-related terminal functions. However, it would be desirable to choose packet sizes to reduce as much as possible to the relative overhead caused by the need for the leader padding on each packet that is necessary for privacy-device synchronization.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

18 19 REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: ESD-TR-78-203

2. GOVT ACCESSION NO.

3. RECIPIENT'S CATALOG NUMBER

4. TITLE (and Subtitle): 6 Information Processing Techniques Program, Volume II, Wideband Integrated Voice/Data Technology,

5. TYPE OF REPORT & PERIOD COVERED: 9 Semiannual Technical Summary, 1 October 1977 – 31 March 1978,

6. PERFORMING ORG. REPORT NUMBER

7. AUTHOR(s): 10 Bernard Gold

8. CONTRACT OR GRANT NUMBER(s): 15 F19628-78-C-0002, ARPA Order-2929

9. PERFORMING ORGANIZATION NAME AND ADDRESS: Lincoln Laboratory, M.I.T., P.O. Box 73, Lexington, MA 02173 207 650

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: ARPA Order 2929, Program Element No. 62708E, Project No. 8T10

11. CONTROLLING OFFICE NAME AND ADDRESS: Defense Advanced Research Projects Agency, 1400 Wilson Boulevard, Arlington, VA 22209

12. REPORT DATE: 11 31 March 1978

13. NUMBER OF PAGES: 40

14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office): Electronic Systems Division, Hanscom AFB, Bedford, MA 01731

15. SECURITY CLASS. (of this report): Unclassified

15a. DECLASSIFICATION DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

12 37 p.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

Volume I is ESD-TR-78-71

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

information processing techniques
wideband integrated voice/data technology
digital voice communications
packet speech
network speech terminal
C^2 links
time-varying communications
wideband satellite network
speech concentration
adaptive networking

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This report describes work performed on the Wideband Integrated Voice/Data Technology program sponsored by the Information Processing Techniques Office of the Defense Advanced Research Projects Agency during the period 1 October 1977 through 31 March 1978.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

207 650